Cluster-based hydrological analysis and prediction of long-term climate change and drought discharge

TOSHIHARU KOJIRI

Water Resources Research Center, DPRI, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

e-mail: tkojiri@wrcn2.dpri.kyoto-u.ac.jp

Abstract Partly due to climate change, natural disasters such as flood inundation or drought have recently caused heavy damage in many places in the world. Estimation and prediction of the present climate situation will provide significant information to reduce these damages. In this paper, pattern classification methodologies are introduced to analyse global warming effects and spatial relationships of considered meteorological events from a global viewpoint. Furthermore, the advanced AI technology of neural networks is applied to predict regional and local precipitation magnitudes and features.

INTRODUCTION

Recently, natural disasters of flood inundation and drought situations were thought to be caused by long-term climate change, that is, changing meteorological and hydrological conditions from a global viewpoint. Moreover, the possibility of such events occurring must be evaluated with respect to not only the total volume but also the feature pattern. Severe storm and drought events are found in unusual meteorological situations showing unexpected fluctuations that may differ strongly even if the volumes are the same.

The distributions of hydrological elements such as precipitation or air temperature can be classified according to the peculiar feature patterns of widely monitored data provided by the weather service. In this paper, spatial and temporal characteristics of air pressure, air temperature and sea surface temperature are classified through pattern classification methods allowing for extraordinary events of precipitation shortage to be analysed with respect to long-term climate change.

Firstly, meteorological characteristics are extracted through the ISODATA pattern recognition concept for the Northern Hemisphere. After defining the objective function for classification, the optimal number of patterns is decided by AIC (Akaike Information Criteria) technology. The correlation between classified elements is evaluated using a fitness concept. Arranging the classified results in time, global warming is shown from the aspects of both the average and feature patterns of the air temperature. The occurrence probability of severe droughts is found by summing the occurrence probability of each considered event belonging to a certain pattern class multiplied by the pattern occurrence probability.

Secondly, the precipitation distribution in Japan is estimated by analysing the correlated elements using fuzzy inference. Meteorologically specified situations such as existence of cold air over Tibet or appearance of El Niño off-coast of Peru are related numerically with fuzzy logic to other spatial characteristics. The future precipitation is predicted with a neural network.

CALCULATION PROCEDURE OF PATTERN CLASSIFICATION

Temporal and spatial images with meteorological or hydrological distributions can be treated as a matrix feature with components of individually monitored data (Toriwaki, 1992). Unny *et al.* (1981) proposed pattern classification for hydrographs to extract their hydrological characteristics. Kojiri *et al.* (1992) estimated the drought feasibility of classified patterns. Herein, the ISODATA method (Tou & Gonzalez, 1974) is introduced for establishing the pattern classification. When the number of clusters is different from the designed one, the criteria for separation/incorporation in the ISODATA method should be changed, or the temporal centre should be added/subtracted at random.

An objective function (*OF*) has to be defined according to the classification purpose. To handle precipitation shortage, the variance around the cluster centre is formulated as follows:

$$OFa(i,m) = \sqrt{\sum_j [X(j,i) - CL(j,m)]^2} \quad i \in m \tag{1}$$

where j denotes the time (month) and summation is taken among the sample data i belonging to the cluster centre, m. In the case of spatial air pressure, the following normalized function is proposed to introduce the equivalence of spatially distributed data:

$$OFb(i,m) = \max_{u,v} \frac{|X(u,v,i) - CL(u,v,m)|}{CL(u,v,m)} \quad i \in m \tag{2}$$

where u and v denote the grid number in longitude and latitude, respectively.

CLASSIFICATION OF SEQUENTIAL HYDROLOGICAL EVENTS

Optimal number and correlation of clusters

After combining the classification process with the objective function (equation (1)), the optimal number of cluster centres should be decided. The distances in the objective function are assumed to obey the log normal distribution with mean $\lambda(m)$ and variance $\varsigma(m)^2$ around the cluster centre. Then the likelihood value in the mth cluster is represented as follows:

$$LL(m) = \sum_i \left\{ \ln[OF(i,m)] + \ln[2\pi\varsigma^2(m)]/2 + [((\ln OF(i,m) - \lambda(m))/\varsigma(m))^2]/2 \right\} \tag{3}$$

The optimal number is obtained with the following minimization:

$$AIC(m) = -2\sum_{m} LL(m) + 4 \rightarrow \min \tag{4}$$

As the classified feature patterns (cluster centres) have their own attributes as functions of particular factors such as whole shape, peak value and so on, the correlation is analysed with the concept of fitness (Suzuki, 1973). For instance, the factor A has a strong relationship with the factor B if the following condition is satisfied:

$$N_{11} / N_{\bullet 1} > N_{12} / N_{\bullet 2} \tag{5}$$

where N_{ij} is the sample number under the joint occurrence condition of pattern i in factor A and pattern j in factor B, $N_{\cdot i}$ is the sample number under the occurrence condition of pattern i in factor A and so on. The above equation can be extended to the case of multi-factors.

Occurrence probability of feature patterns

As, traditionally, the occurrence probability has been expressed using the total volume of considered event such as yearly precipitation or discharge amount for water shortage, extraordinary features have not been evaluated properly for the reason of single aspect of total volume. A new approach to estimate the occurrence probability is proposed by combining the traditional method with pattern recognition. The occurrence probability of a considered event is represented with the following equations:

$$PX = \sum PP(m) \tag{6}$$

$$PP(m) = PC(m) \cdot PD(X|m) \tag{7}$$

where $PP(m)$ is the probability of an event belonging to pattern m, $PC(m)$ is the occurrence probability of cluster m and $PD(X|m)$ is the conditional occurrence probability of X given cluster m. $N(m)$ is the sample number in cluster m and NT is the total number of data. Assuming the lognormal distribution, both functions are formulated as follows:

$$PC(m) = N(m) / NT \tag{8}$$

$$PD(X|m) = \frac{1}{\sqrt{2\pi}\varsigma(m)} \int_{OF\bullet(X,m)}^{\infty} \frac{1}{\tau} \exp\left[-\frac{1}{2} \left\{ \frac{\ln(\tau) - \lambda(m)}{\varsigma(m)} \right\}^2 \right] d\tau \tag{9}$$

CLASSIFICATION OF SPATIAL DATA

Similarity of climate characteristics

The meteorological conditions leading to precipitation shortage in Japan is estimated by two factors, namely height distribution of air pressure and distribution of sea surface temperature. The height of 500 hPa air pressure is provided at each 10° for both latitude and longitude in the Northern Hemisphere. A first characteristic of height distribution of air pressure is extracted using equation (2). A second criterion is the coefficient of shape, which is the mean value between compactness measures of 5200 and 5600 m height as follows:

$$CHi = PHi^2 /(4\pi AHi) \tag{10}$$

where AHi is the distribution area of the considered height of air pressure, and PHi is the periphery according to the area AHi. The third one is the skewness expressed by the distance between the lowest height $APMi$ of the considered air pressure and the farthest point $APVi$ as follows:

$$TVi = \max|APVi - APMi| \tag{11}$$

These three objective values are integrated through averaging in a fuzzy set as follows:

$$OFI(i,m) = [\mathrm{fm}O(OF\cdot) + \mathrm{fm}C(CHi) + \mathrm{fm}T(TVi)]/3 \tag{12}$$

where fm(·) means the fuzzy membership function.

Sea surface temperature is also an important factor for recognizing the meteorological characteristics. Data are provided at every two degrees of latitude and longitude at the global scale. Introducing the concept of image processing, the objective function consists of four sub-functions: The first criterion $OFSi$ is the same function as given for air pressure by equation (2) and used to discriminate the fluctuation of surface temperature. The second one is to obtain the location of the maximum sea surface temperature and is described as the gradient from Japan as follows:

$$SSi = yi / xi \tag{13}$$

where yi and xi are the converted distances from Japan in sample data i of latitude and longitude of the maximum temperature. The third criterion is to evaluate the specified area and is described by the difference between width and length of high temperature zones where the sea surface temperature is higher than a threshold value as follows:

$$NSi = |TW\max i - TH\max i| \tag{14}$$

where $TW\mathrm{max}i$ means the longest longitude width of the higher sea surface temperature part and $TH\mathrm{max}i$ is the longest latitude length of same part. The last criterion is used to evaluate the impact of El Niño which seems to cause unusual weather scenarios such as cold summers or hot winters and is given as follows:

$$ELSi = \sqrt{\sum_{u,v} [X(u,v,i) - CL(u,v,m)]^2} \tag{15}$$

The analysed area in equation (15) is from 40°N to 10°S and from 90°W to 140° off-coast of Peru. Assuming the fuzzy membership function to include all criteria, the pattern classification is established on basis of the integrated similarity of sea surface temperature obtained through the hierarchical process shown in Fig. 1.

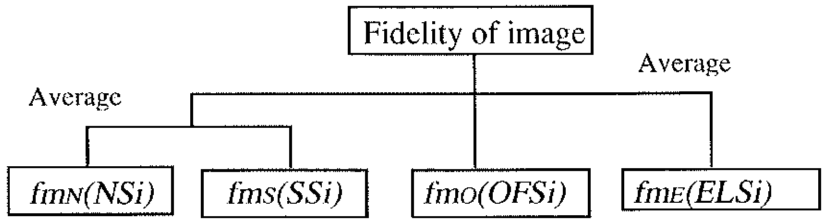

Fig. 1 Calculation procedure of pattern classification for sea surface temperature.

Prediction of precipitation at gauging stations

In order to prepare counter-measures for water shortage, several month-ahead predictions of precipitation are requested in the considered regions. A two-step approach is applied to predict the precipitation at the gauging stations: first prediction is made at the regional scale, and subsequently the prediction is at an arbitrary gauging station. Temporal and spatial precipitation data in Japan are classified by the ISODATA method with the normalized objective function as follows:

$$OFp(i,m) = \sqrt{\sum_j \{[X(j,i) - CL(j,m)]/CL(j,m)\}^2} \tag{16}$$

Air temperature data are classified with the same objective function as precipitation, i.e. equation (16). The correlation between patterns of air pressure, sea surface temperature, air temperature and precipitation are evaluated considering fitness property. A fuzzy set is introduced to take account of the time hysteresis of events by employing *IF-THEN* forms as follows:

$$\begin{aligned} &IF(\text{the monitored sequence is pattern CPSD}(m)) \\ &THEN(\text{future feature should follow CPSD}(m) \text{ with calculated fidelity}) \end{aligned} \tag{17}$$

where CPSD is the sequence of considered events according to the classified pattern in the sample data. The fidelity in equation (17) is represented by an antecedent fuzzy grade (*Wij*) similar to the integrated objective function equation (12) by averaging between fuzzy membership grades of monitored sequence against the classified pattern. The output *FP*, which denotes the predicted precipitation sequence, is inferred by the fuzzy height method as follows:

$$FP = \frac{\sum W_{ij} CPSD_{ij}}{\sum W_{ij}} \tag{18}$$

The precipitation at main gauging stations is predicted with a neural network of the back propagation type. The predicted air pressure, sea surface temperature, air temperature and present precipitation are used as input and one hidden layer is applied. Subsequently, the predicted precipitation data at arbitrary points or river basins can be established with similar neural networks for evaluation of reservoir operation, water saving and other activities related to water shortage.

APPLICATION OF PROPOSED METHODOLOGIES

Climate change and occurrence probability

To verify the proposed methodologies, real data and a real river catchment are applied. The data period of 106 years at the Regional Meteorology Station at Nagoya is employed for classification of precipitation and air temperature. Setting the proper values for classification, the ISODATA method was applied to obtain the designed cluster number. The AIC sequence against cluster number yielded the optimal number of 12 as the minimum point for precipitation and 10 for air temperature. Figure 2 shows the classified precipitation sequence represented by the average.

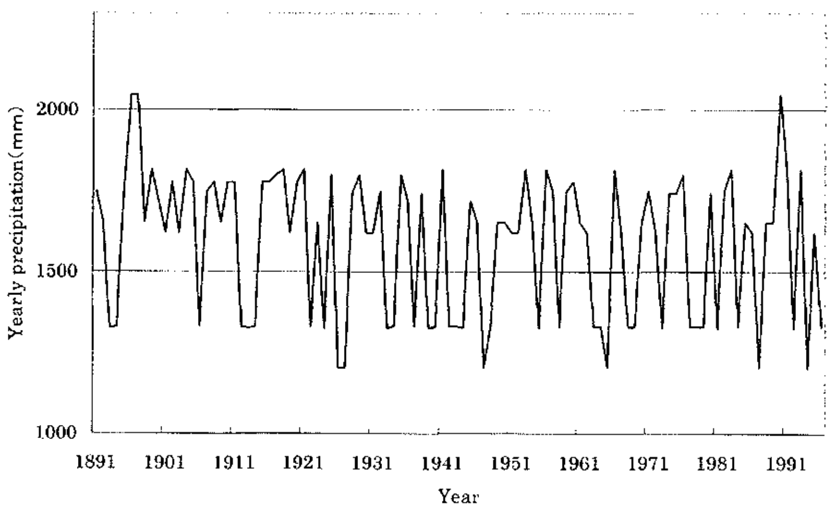

Fig. 2 Classified result of height distribution of air pressure in January.

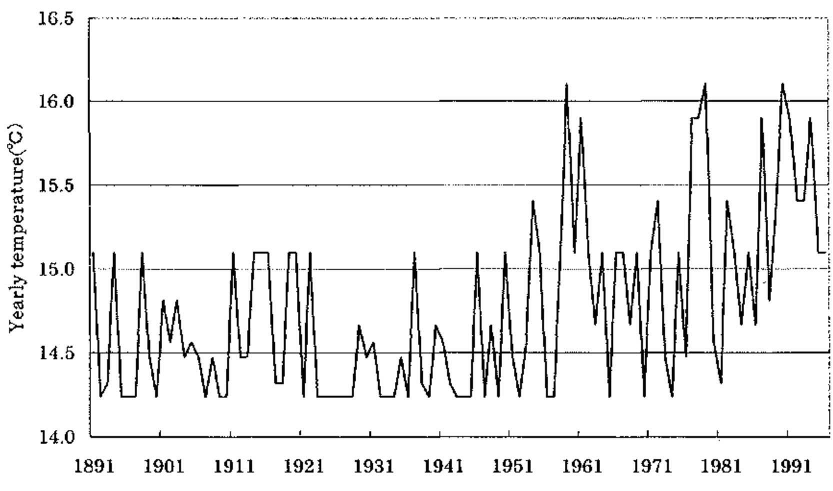

Fig. 3 Classified result of sea surface temperature in February.

Though the annual precipitation seems to be decreasing from a long-term aspect, the distinguished feature pattern due to global warming was not clearly obtained and the rare pattern happened to occur in the recent decade. On the other hand, the air temperature seems to be increasing. The typical feature patterns with higher average, which seem to be caused by global warming, have happened frequently in these years as shown in Fig. 3.

Japan suffered from a severe drought in 1994 and the annual precipitation was 1061 mm. The return period was 96.2 years determined by the traditional probability approach and 73.5 years using the proposed method. It means that this drought pattern may occur frequently under today's climate conditions.

Spatial classification

The period of arranged data was 28 years from 1960 to 1987 for air pressure, sea surface temperature at the global scale, and air temperature and precipitation at main gauging stations in Japan. Figure 4 shows classified air pressure distributions in January where a number of classes of six was set in advance for calculation convenience. Each pattern expresses the different feature of air pressure height around the North Pole. Figure 5 shows the classified patterns of sea surface temperature in January and the obvious difference around the Equator except for the effect of El Niño.

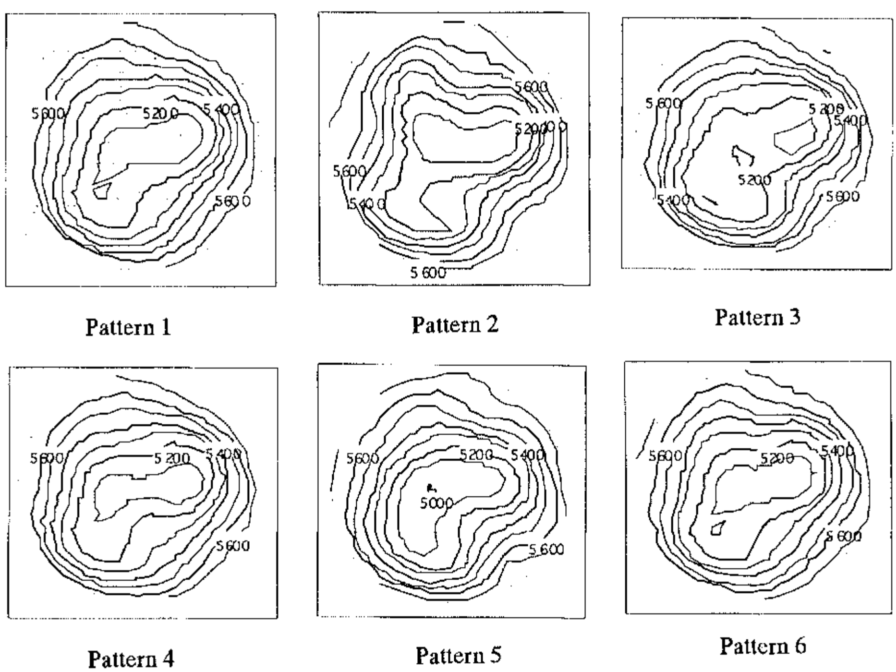

Fig. 4 Classified result of temperature patterns represented with average.

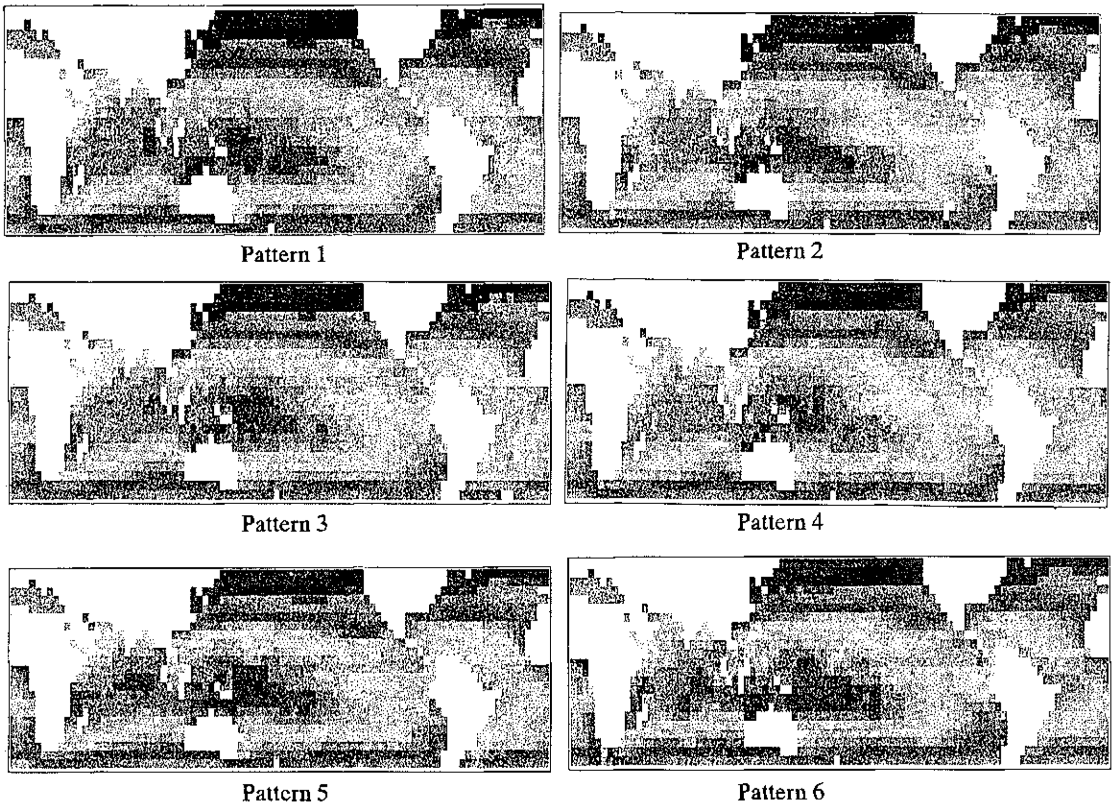

Fig. 5 Classified result of precipitation patterns represented with average.

Prediction of precipitation at considered stations

Figure 6 shows the comparison of predicted and observed precipitation at main gauging stations in Japan. At some stations, where the precipitation has a strong relationship to the applied components, good prediction results were obtained, while at other stations less favourable results were obtained because of a weak relationship. The predicted precipitation at Hikone, not belonging to the set of main stations, was 967 mm against 1056 mm of monitored precipitation. Precipitation in 1994 was predicted with a neural network with fixed parameters, which were identified using 1986. As the neural network can provide the same accuracy in real-time predictions as in the identification process, an effective drought control can be established by using the proposed system.

CONCLUSIONS

In this paper, the occurrence possibility and prediction of hydrological events through pattern classification have been discussed. To sum up, the following results were obtained:

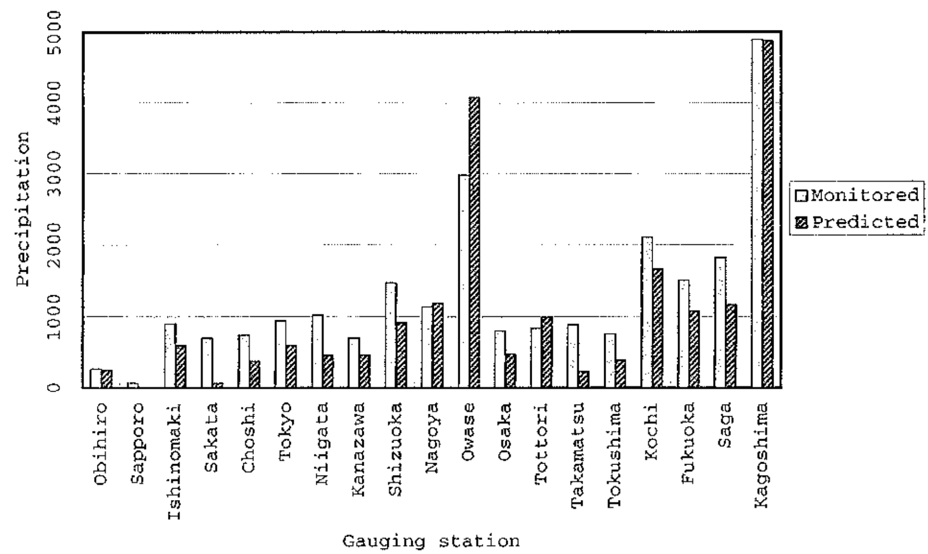

Fig. 6 Predicted result of precipitation at main gauging stations in Japan.

(a) Formulation of pattern classification in time and space Pattern classification procedures were formulated with the ISODATA method and the optimal number of classified cluster centres was decided by use of AIC. Seasonal classification appeared more effective than yearly cluster patterns to recognize the typical characteristics of feature patterns. The correlation between different events was analysed using the concept of fitness.

(b) Extraction of temporal characteristics of climate change Classified temperature sequences verified the recent gradient phenomena and the patterns of climate change. It appeared difficult to incorporate the precipitation sequence for correlation in relation to global warming, though the classification process was easily established.

(c) Extraction of spatial characteristics of hydrological events Introducing image processing and fuzzy sets, the spatial data could be classified. The distributions of air pressure, air temperature and sea surface temperature were classified from the viewpoint of water shortage.

(d) Prediction of precipitation in a regional area The regional three-month-ahead precipitation was inferred through fuzzy inference theory and then the precipitation at the considered gauging station was predicted using a neural network. Drought prevention and effective counter-measures might be established by using the predicted information.

REFERENCES

Kojiri, T., Unny, T. E. & Panu, U. S. (1992) Estimation and prediction of drought by using pattern recognition technique. In: *Proc. 6th IAHR Int. Symp. on Stochastic Hydraulics*, 711–718.

Suzuki, E. (1973) *Meteorological Statistics*, 254–261. Jinti Publishing Co., Japan.

Toriwaki, J. (1992) *Pattern Recognition and Image Processing*, 56–65. Asakura Publishing, Japan.

Tou, J. T. & Gonzalez, R. C. (1974) Pattern recognition principle. In: *Applied Mathematics and Computation*, no. 7, 75–109. Addison Wesley.

Unny, T. E., Panu, U. S., Macinees, C. D. & Wong, A. K. K. (1981) Pattern analysis and synthesis of time-dependent hydrological data. *Adv. Hydrosci.* **12**, 222–244.